An Overview of Multilinear Algebra and Tensor Decompositions

ARCC Tensor Decomposition Workshop

Carla D. Moravitz Martin
Center for Applied Mathematics
Cornell University

July 19, 2004

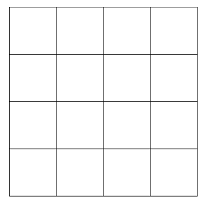
Outline

- Representing Tensors
- Tensor Rank Concepts
- Algorithms and SVD Generalizations

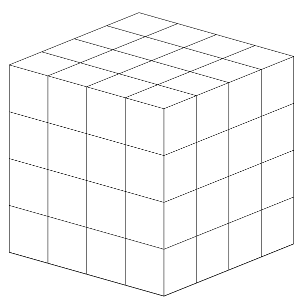
Representing Tensors
(a.k.a. multiway arrays)

Order

- Second-order tensor $A = (a_{ij}) \in \mathbb{R}^{n_1 \times n_2}$



- Third-order tensor $\mathcal{A} = (a_{ijk}) \in \mathbb{R}^{n_1 \times n_2 \times n_3}$



- p^{th}-order tensor $\mathcal{A} = (a_{i_1 i_2 \ldots i_p}) \in \mathbb{R}^{n_1 \times \cdots \times n_p}$

Some Relations to Linear Algebra

- Tensors as matrices

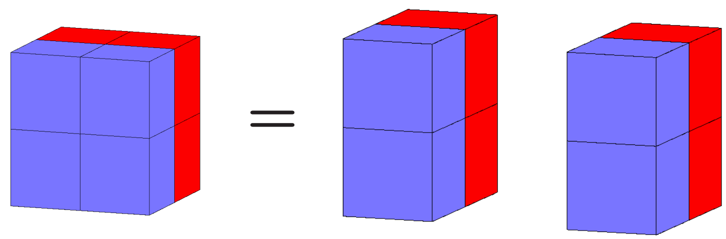
- Tensors as vectors

- Norms, inner products, outer products

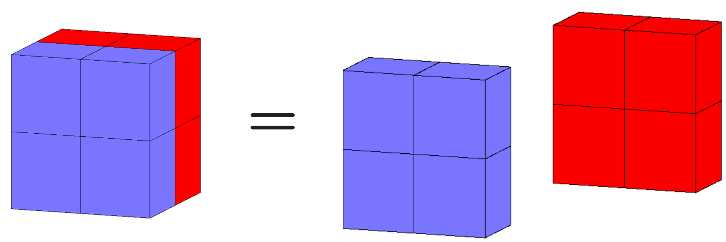
Turning Tensors into Matrices

Three ways to cut a “cube”:

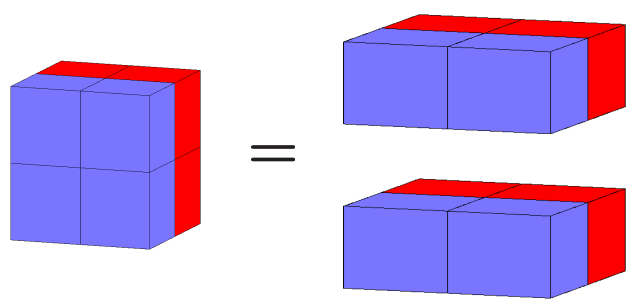
1. Left-right



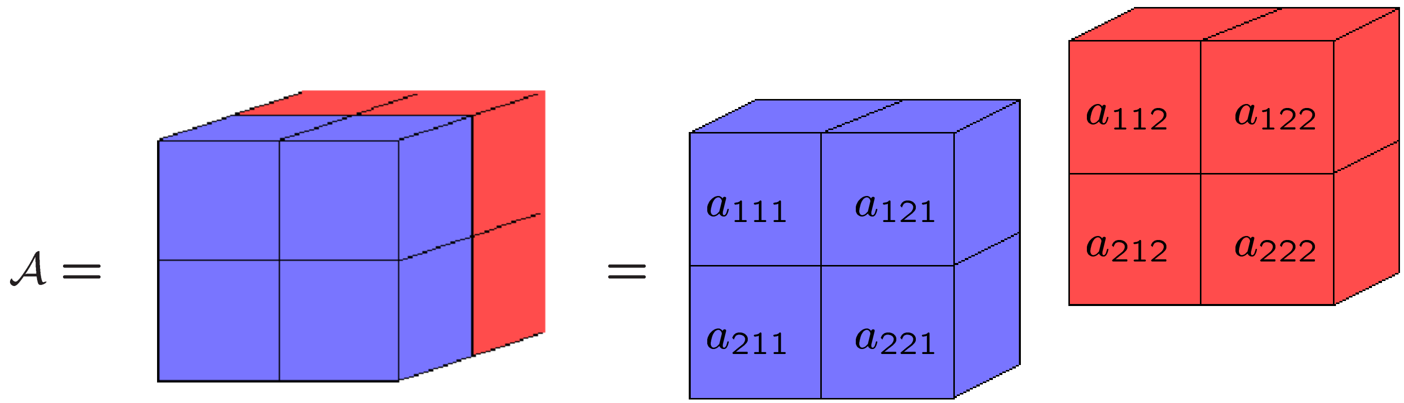
2. Front-back



3. Top-bottom



Unfolding Matrices*



$$A_{(1)} = \left[\begin{array}{cc|cc} a_{111} & a_{112} & a_{121} & a_{122} \\ a_{211} & a_{212} & a_{221} & a_{222} \end{array}\right] \quad \text{“sides”}$$

$$A_{(2)} = \left[\begin{array}{cc|cc} a_{111} & a_{211} & a_{112} & a_{212} \\ a_{121} & a_{221} & a_{122} & a_{222} \end{array}\right] \quad \text{“front-back” [transposed]}$$

$$A_{(3)} = \left[\begin{array}{cc|cc} a_{111} & a_{121} & a_{211} & a_{221} \\ a_{112} & a_{122} & a_{212} & a_{222} \end{array}\right] \quad \text{“top-bottom” [transposed]}$$

*De Lathauwer, De Moor, Vandewalle (2000)

The `vec` and `reshape` Operators

$$z \in \mathbb{R}^{mn} \Rightarrow \texttt{reshape}(z, m, n) \in \mathbb{R}^{m \times n}$$

Example: $m = 3$, $n = 5$

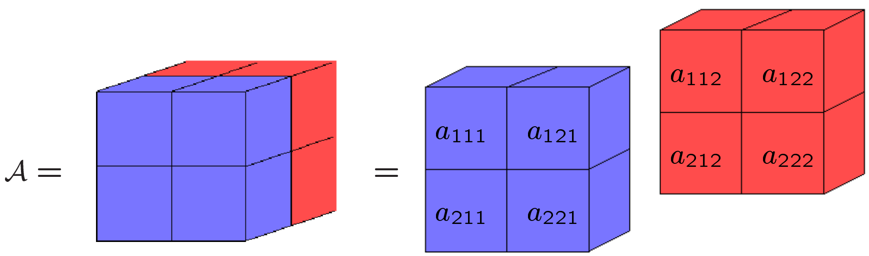
$$\texttt{reshape}(z, 3, 5) = \begin{bmatrix} z_1 & z_4 & z_7 & z_{10} & z_{13} \\ z_2 & z_5 & z_8 & z_{11} & z_{14} \\ z_3 & z_6 & z_9 & z_{12} & z_{15} \end{bmatrix}$$

$$Z \in \mathbb{R}^{m \times n} \Rightarrow \texttt{vec}(Z) = \texttt{reshape}(Z, mn, 1) \begin{bmatrix} Z(:,1) \\ \vdots \\ Z(:,n) \end{bmatrix} \in \mathbb{R}^{mn}$$

Turning Tensors into Vectors

$$\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3} \quad \Rightarrow \quad \mathtt{vec}(\mathcal{A}) \in \mathbb{R}^{n_1 n_2 n_3}$$

Example: For $n = 2$,



$$\mathtt{vec}(\mathcal{A}) = \begin{bmatrix} a_{111} \\ a_{211} \\ a_{121} \\ a_{221} \\ a_{112} \\ a_{212} \\ a_{122} \\ a_{222} \end{bmatrix}$$

Unfoldings in Matlab

The s-th unfolding matrix of an order-p tensor can be expressed in terms of `vec` and `reshape`:

Example:

$$\text{“}A_{(s)} = \texttt{reshape}(\texttt{reshape}(\texttt{vec}(\mathcal{A}), n^{p-s}, n^s)^T, n, n^{p-1})\text{”}$$

(can be adjusted for unequal dimensions)

Example: $n_1 \times n_2 \times n_3$ third-order tensor:

$$\text{“}A_{(2)} = \texttt{reshape}(\texttt{reshape}(\texttt{vec}(\mathcal{A}), n_1, n_2 n_3)^T, n_2, n_3 n_1)\text{”}$$

Caution about notation

The explicit combination of `vec` and `reshape` in

$$\text{``}A_{(s)} = \texttt{reshape}(\texttt{reshape}(\texttt{vec}(\mathcal{A}), n^{p-s}, n^s)^T, n, n^{p-1})\text{ ``}$$

depends on how the unfoldings are defined.

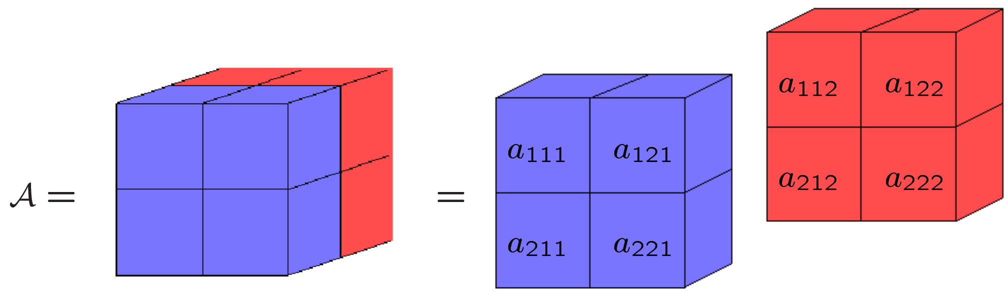
Example: Using DDV* unfoldings,

$$\texttt{vec}(\mathcal{A}) = \begin{bmatrix} a_{111} \\ a_{211} \\ a_{121} \\ a_{221} \\ a_{112} \\ a_{212} \\ a_{122} \\ a_{222} \end{bmatrix} \xrightarrow{permutation} \begin{bmatrix} a_{111} \\ a_{112} \\ a_{121} \\ a_{122} \\ a_{211} \\ a_{212} \\ a_{221} \\ a_{222} \end{bmatrix}$$

*De Lathauwer, De Moor, Vandewalle (2000)

Subtensors in Matlab

If



then

$$A(:,:,1) = \begin{bmatrix} a_{111} & a_{121} \\ a_{211} & a_{221} \end{bmatrix}.$$

However,

$$A(1,:,:) \neq \begin{bmatrix} a_{111} & a_{112} \\ a_{121} & a_{122} \end{bmatrix}$$

but rather the $1 \times 2 \times 2$ tensor: $\begin{bmatrix} a_{111} & a_{121} \end{bmatrix} \quad \begin{bmatrix} a_{112} & a_{122} \end{bmatrix}$

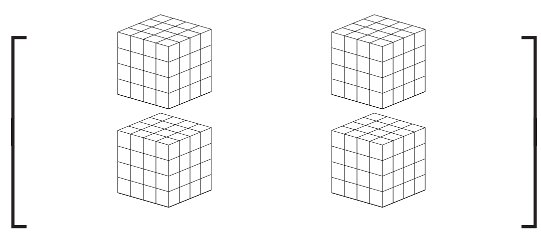
Block Representations

A matrix of scalars, $A = (a_{ij})$, can be regarded as a "matrix with matrix entries" (*block matrix*)

A tensor of scalars, $\mathcal{A} = (a_{i_{n_1} \ldots i_{n_p}})$, can be regarded as a "matrix with tensor entries"

Example: Order-5 tensor:

$$\begin{bmatrix} \mathcal{A}_{11} & \mathcal{A}_{12} \\ \mathcal{A}_{21} & \mathcal{A}_{22} \end{bmatrix} =$$



"second-order with third-order entries"

Norms and Inner Products

If $\mathcal{A}, \mathcal{B} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ then the *inner product* is

$$\begin{aligned} < \mathcal{A}, \mathcal{B} > &= \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \sum_{k=1}^{n_3} a_{ijk} \cdot b_{ijk} \\ &= \texttt{vec}(\mathcal{A})^T \cdot \texttt{vec}(\mathcal{B}) \end{aligned}$$

A *Frobenius norm*, $\|\mathcal{A}\|$, is

$$\|\mathcal{A}\| = \|\mathcal{A}\|_F = \sqrt{< \mathcal{A}, \mathcal{A} >}$$

Other norms?

Outer Product and Rank-1 Tensors

If $x \in \mathbb{R}^{n_1}, y \in \mathbb{R}^{n_2}$ then the *outer product*, yx^T, is a rank-1 matrix.

Note

$$\texttt{vec}(yx^T) \quad \Leftrightarrow \quad x \otimes y$$

More generally, if x, y, z are vectors,

$$x \otimes y \otimes z$$

is a *rank-1* tensor.

Sums of Rank-One Matrices

If $A = U\Sigma V^T$, $U = \begin{bmatrix} u_1 & \dots & u_n \end{bmatrix}, V = \begin{bmatrix} v_1 & \dots & v_n \end{bmatrix}$, then

$$A = \sum_{i=1}^{n} \sum_{j=1}^{n} \sigma_{ij} u_i v_j^T$$

that is,

$$\begin{aligned} \texttt{vec}(A) &= \sum_{i=1}^{n} \sum_{j=1}^{n} \sigma_{ij}(v_j \otimes u_i) \\ &= (V \otimes U) \cdot \texttt{vec}(\Sigma) \end{aligned}$$

Sums of Rank-One Tensors

$$\begin{aligned} \mathtt{vec}(\mathcal{A}) &= \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} \sigma_{ijk} (w_k \otimes v_j \otimes u_i) \\ &= (W \otimes V \otimes U) \cdot \mathtt{vec}(\Sigma) \end{aligned}$$

where

$$\begin{aligned} U &= \begin{bmatrix} u_1 & \ldots & u_n \end{bmatrix} \\ V &= \begin{bmatrix} v_1 & \ldots & v_n \end{bmatrix} \\ W &= \begin{bmatrix} w_1 & \ldots & w_n \end{bmatrix} \end{aligned}$$

Repeated Change of Basis

If

$$\begin{aligned} \texttt{vec}(\mathcal{A}) &= \sum_i \sum_j \sum_k b_{ijk}(w_k \otimes v_j \otimes u_i) \\ \texttt{vec}(\mathcal{B}) &= \sum_i \sum_j \sum_k c_{ijk}(\widehat{w}_k \otimes \widehat{v}_j \otimes \widehat{u}_i) \end{aligned}$$

Then

$$\begin{aligned} \texttt{vec}(\mathcal{A}) &= \sum_i \sum_j \sum_k c_{ijk}(W\widehat{w}_k \otimes V\widehat{v}_j \otimes U\widehat{u}_i) \\ &= (W\widehat{W} \otimes V\widehat{V} \otimes U\widehat{U}) \cdot \texttt{vec}(\mathcal{C}) \end{aligned}$$

Connections between the unfoldings of $\mathcal{A}$ and Σ

$$\mathtt{vec}(\mathcal{A}) = (W \otimes V \otimes U) \cdot \mathtt{vec}(\Sigma)$$

$$\Updownarrow$$

$$\begin{aligned}
A_{(1)} &= U \Sigma_{(1)} (V \otimes W)^T \\
A_{(2)} &= V \Sigma_{(2)} (W \otimes U)^T \\
A_{(3)} &= W \Sigma_{(3)} (U \otimes V)^T
\end{aligned}$$

n-mode Products

Let $\mathcal{A} \in \mathbb{R}^{m \times n \times p}$

		matrix $\times$ matrix [unfolding]	matrix $\times$ vector [vec]
$B_1 \in \mathbb{R}^{q \times m}$	"$\mathcal{A} \times_1 B_1$" $(q \times n \times p)$	$B_1 \cdot A_{(1)}$	$(I \otimes B_1) \cdot \texttt{vec}(A_{(1)})$
$B_2 \in \mathbb{R}^{q \times n}$	"$\mathcal{A} \times_2 B_2$" $(m \times q \times p)$	$B_2 \cdot A_{(2)}$	$(I \otimes B_2) \cdot \texttt{vec}(A_{(2)})$
$B_3 \in \mathbb{R}^{q \times p}$	"$\mathcal{A} \times_3 B_3$" $(m \times n \times q)$	$B_3 \cdot A_{(3)}$	$(I \otimes B_3) \cdot \texttt{vec}(A_{(3)})$

Tensor Rank Concepts

General Tensor Rank

Tensor rank of $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ is the minimum number of rank-1 tensors that sum to $\mathcal{A}$ in linear combination.

If a tensor $\mathcal{A}$ has a minimal representation as:

$$\mathtt{vec}(\mathcal{A}) = \sum_{i=1}^{r_1} \sum_{j=1}^{r_2} \sum_{k=1}^{r_3} \sigma_{ijk}(w_k \otimes v_j \otimes u_i)$$

then $\mathrm{rank}(\mathcal{A}) = r_1 r_2 r_3$

Why is Tensor Rank Important?

- Enables data compression

- Identifies dependencies in data

Applications, multilinear algebra theory, and computational realities all have "something to say" about the tensor rank issue.

Eight Facts about Tensor Ranks

1. Minimum tensor representation not necessarily orthogonal

2. Different orthogonality requirements result in different minimal representations (Kolda, 2001)

3. Ranks in different dimensions not always equal
   (De Lathauwer, De Moor, Vandewalle, 2000)

4. Tensors can't always be "diagonalized"

Eight Facts about Tensor Ranks (continued)

5. Maximum tensor rank unknown in general

6. No known method to compute the "minimum" tensor representation

7. k successive rank-1 approximations to tensors do not necessarily result in the best rank-k approximation
   (Zhang and Golub, 2001; Kolda, 2001,2003)

8. Set of rank-deficient tensors has positive volume (Kruskal, 1989)

Rank Analogy with Matrices

Matrices $A \in \mathbb{R}^{n\times n}$

$$\texttt{vec}(A) = \sum_{i=1}^{n}\sum_{j=1}^{n} \sigma_{ij}(v_j \otimes u_i)$$

$$\Downarrow$$

[orthogonal decomposition]

$$\texttt{vec}(A) = \sum_{i=1}^{r} \sigma_i(v_i \otimes u_i)$$

where $\text{rank}(A) = r$

TRUE

Tensors $\mathcal{A} \in \mathbb{R}^{n\times n\times n}$

$$\texttt{vec}(\mathcal{A}) = \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n} \sigma_{ijk}(w_k \otimes v_j \otimes u_i)$$

$$\Downarrow$$

[orthogonal decomposition]

$$\texttt{vec}(\mathcal{A}) = \sum_{i=1}^{r} \sigma_i(w_i \otimes v_i \otimes u_i)$$

where $\text{rank}(\mathcal{A}) = r$

FALSE (most of the time)

Notions of Orthogonality (Kolda, 2001)

$$\mathtt{vec}(\mathcal{A}) = \sum_{i=1}^{n} \sigma_i (w_i \otimes v_i \otimes u_i)$$

1. *Complete Orthogonal Decomposition*: for $i, j = 1, \ldots, n$

$$u_i \perp u_j \quad \text{and} \quad v_i \perp v_j \quad \text{and} \quad w_i \perp w_j$$

2. *Strong Orthogonal Decomposition*: for $i, j = 1, \ldots, n$

$$\begin{array}{rcl} u_i \perp u_j & \text{or} & u_i = \pm u_j \\ v_i \perp v_j & \text{or} & v_i = \pm v_j \\ w_i \perp w_j & \text{or} & w_i = \pm w_j \end{array}$$

3. *Orthogonal Decomposition*: for $i, j = 1, \ldots, n$

$$u_i \perp u_j \quad \text{or} \quad v_i \perp v_j \quad \text{or} \quad w_i \perp w_j$$

Orthogonality and Rank *

If the rank-1 tensors $(w_i \otimes v_i \otimes u_i)$ are $\left\{ \begin{array}{c} \textit{orthogonal} \\ \textit{strongly orthogonal} \end{array} \right\}$ in

$$\texttt{vec}(\mathcal{A}) = \sum_{i=1}^{r} \sigma_i (w_i \otimes v_i \otimes u_i)$$

and r is minimal, then the $\left\{ \begin{array}{c} \textit{orthogonal} \\ \textit{strong orthogonal} \end{array} \right\}$ *rank* of $\mathcal{A}$ is r.

*Kolda, 2001

Orthogonality and Rank (continued)

If the rank-1 tensors $(w_k \otimes v_j \otimes u_i)$ are *completely orthogonal* in

$$\texttt{vec}(\mathcal{A}) = \sum_{i=1}^{r} \sum_{j=1}^{r} \sum_{k=1}^{r} \sigma_{ijk}(w_k \otimes v_j \otimes u_i)$$

and r is minimal, then the *combinatorial orthogonal rank* of $\mathcal{A}$ is r.

More Rank Concepts*

For an order-p tensor, compute rank of each unfolding matrix:

$$n\text{-rank of } \mathcal{A} = \text{rank}(A_{(n)})$$

Relationship to tensor rank:

$$n\text{-rank} \leq \text{rank}$$

*De Lathauwer, De Moor, Vandewalle, 2000

Matrix Rank vs. Tensor Ranks

- *Matrices*: rank, orthogonal rank, strong orthogonal rank, and combinatorial rank are all equal

- *Tensors*: different ranks are not necessarily equal and the associated decompositions are not unique

Matrix Rank vs. Tensor Ranks (continued)

- *Matrices*: the n-ranks correspond to the column and row rank of the matrix and hence are equal

- *Tensors*: the different n-ranks are not necessarily equal and even if they are, they do not necessarily equal the tensor rank

Diagonalizeable Tensors

Suppose $\mathcal{A}$ can be written as a *completely* orthogonal decomposition:

$$\texttt{vec}(\mathcal{A}) = \sum_{i=1}^{n} \sigma_i (w_i \otimes v_i \otimes u_i)$$

Then

- All ranks with different orthogonality constraints are equal
- k successive rank-1 approximations compute the best rank-k approximation (Zhang and Golub, 2001)

Other Special Structures to Explore

- Supersymmetric tensors
- $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, where $n_1, n_2 >> n_3$ (*thin* tensors)
- Bandedness?
- Toeplitz?
- Block algorithms?

Thinking About Rank

- How does knowledge of rank help in applications?

- How do the different notions of rank and orthogonality help explain correlations in data?

- Can we overcome the rank problem by instead working to “compress” entries of a tensor?

Numerical Rank

Tricky even in matrix case!

$$\texttt{vec}(A) = \sum_{i=1}^{r} \sigma_i (v_i \otimes u_i)$$

where

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r >> \sigma_{r+1} \geq \cdots \geq \sigma_n$$

How to determine r?

Numerical rank problems with tensors?

Algorithms
and
SVD Generalizations

Solution Paradigm: Component-wise Linearization

Problem:

$$\min_{u,v,w} \|a - w \otimes v \otimes u\|_F^2$$

Repeat:

Hold w, v constant, solve for u
Hold w, u constant, solve for v
Hold v, u constant, solve for w

"Alternating Least Squares"

Best Rank-1 Idea

Problem:

$$\min_{u,v,w} \|a - \sigma(w \otimes v \otimes u)\|_F^2 \quad \text{s.t.} \quad \|u\| = \|v\| = \|w\| = 1$$

$$\Updownarrow$$

$$\max_{u,v,w} \; a^T(w \otimes v \otimes u) \quad \text{s.t.} \quad \|u\| = \|v\| = \|w\| = 1$$

Lagrange Multipliers:

$$\begin{aligned} \sigma u &= A_{(1)}(v \otimes w) \\ \sigma v &= A_{(2)}(u \otimes w) \\ \sigma w &= A_{(3)}(u \otimes v) \\ \sigma &= a^T(w \otimes v \otimes u) \end{aligned}$$

Higher-Order Power Method *

Alternating Least Squares:

Solve for σ, u given v, w and iterate:

$$\begin{aligned} \tilde{u} &\leftarrow A_{(1)}(v \otimes w) \\ \sigma &\leftarrow ||\tilde{u}|| \\ u &\leftarrow \frac{\tilde{u}}{\sigma} \end{aligned}$$

*De Lathauwer, De Moor, Vandewalle, 2000

Generalized Rayleigh Quotient Iteration *

Recall the Lagrange equations:

$$
\begin{aligned}
\sigma u &= A_{(1)}(v \otimes w) \\
\sigma v &= A_{(2)}(u \otimes w) \\
\sigma w &= A_{(3)}(u \otimes v) \\
GRQ \equiv \sigma &= a^T(w \otimes v \otimes u)
\end{aligned}
$$

Linearize using Newton's Method and iterate

*Zhang and Golub, 2001

Successive Rank-1 Approximations

Step k:

$$r_k \leftarrow a - \sum_{i=1}^{k} \sigma_i (w_i \otimes v_i \otimes u_i)$$

Solve

$$\max_{u,v,w} \; r_k^T (w \otimes v \otimes u)$$

subject to required orthogonality constraints

Note: Only computes minimal decomposition if tensor is "diagonalizeable" *

*Zhang and Golub, 2001; Kolda, 2001

Existing Algorithms

- CANDECOMP-PARAFAC*
- TUCKER†

Implemented in N-way Toolbox in Matlab‡

*Carroll and Chang, 1970; Harshman, 1970

†Tucker, 1966

‡Andersson and Bro, 2000

CANDECOMP-PARAFAC

Finds general decompositions of the form

$$\texttt{vec}(\mathcal{A}) = \sum_{i=1}^{n} (w_i \otimes v_i \otimes u_i)$$

using ALS on unfolding matrices.

Example:

If V, W fixed, then U is found using LS, where

$$A_{(1)} = U\tilde{I}(V \otimes W)^T \quad \text{and} \quad \tilde{I} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad \text{for } n = 2$$

TUCKER

Finds decompositions of the form

$$\texttt{vec}(\mathcal{A}) = \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} \sigma_{ijk}(w_k \otimes v_j \otimes u_i)$$

using ALS on unfolding matrices (SVD solution at each step)

SVD Generalization (TUCKER3)*

Orthogonal representation involves computing SVDs of the unfolding matrices of $\mathcal{A}$:

$$
\begin{aligned}
A_{(1)} &= U D_1 G_1^T \\
A_{(2)} &= V D_2 G_2^T \\
A_{(3)} &= W D_3 G_3^T
\end{aligned}
\qquad\qquad
\begin{aligned}
\Sigma_{(1)} &= D_1 G_1^T (V \otimes W) \\
\Sigma_{(2)} &= D_2 G_2^T (W \otimes U) \\
\Sigma_{(3)} &= D_3 G_3^T (U \otimes V)
\end{aligned}
$$

Then

$$
\begin{aligned}
A_{(1)} &= U \Sigma_{(1)} (V \otimes W)^T \\
A_{(2)} &= V \Sigma_{(2)} (W \otimes U)^T \\
A_{(3)} &= W \Sigma_{(3)} (U \otimes V)^T
\end{aligned}
$$

*De Lathauwer, De Moor, Vandewalle (2000)

SVD Generalization*

Matrices $A \in \mathbb{R}^{n\times n}$	Tensors $\mathcal{A} \in \mathbb{R}^{n\times n\times n}$
$A = U\Sigma V^T$	$A_{(1)} = U\Sigma_{(1)}(V \otimes W)^T$ $A_{(2)} = V\Sigma_{(2)}(W \otimes U)^T$ $A_{(3)} = W\Sigma_{(3)}(U \otimes V)^T$
$\Sigma = \texttt{diag}(\sigma_1, \ldots, \sigma_n)$ [row/colum vectors mutually orthogonal]	Cuts of Σ mutually orthogonal
$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n \geq 0$	$\|\Sigma(1,:,:)\|_F \geq \cdots \geq \|\Sigma(n,:,:)\|_F \geq 0$ $\|\Sigma(:,1,:)\|_F \geq \cdots \geq \|\Sigma(:,n,:)\|_F \geq 0$ $\|\Sigma(:,:,1)\|_F \geq \cdots \geq \|\Sigma(:,:,n)\|_F \geq 0$

*De Lathauwer, De Moor, Vandewalle (2000)

Conclusions

- Representing tensors with linear algebra tools key!
- How do applications need the different tensor ranks?
- Applications need to drive the algorithms